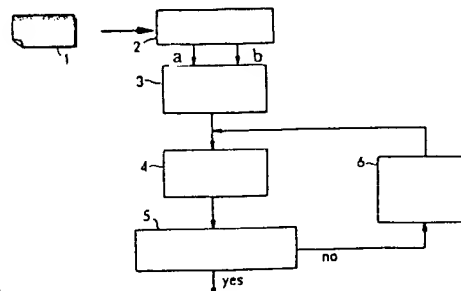


(54) PRE-EDITING ASSISTING PROCESSOR OF MACHINE TRANSLATION DEVICE

(11) 4-153878 (A) (43) 27.5.1992 (19) JP
 (21) Appl. No. 2-280336 (22) 18.10.1990
 (71) FUJITSU LTD (72) JUN IBUKI(4)
 (51) Int. Cl⁵. G06F15/38

PURPOSE: To select a proper equivalent by displaying a sentence which is so rewritten that the structure is made clear as to a predetermined sentence pattern in a source sentence as a result of a sentence analysis and passing the rewritten sentence to a machine translating process.

CONSTITUTION: An analytic part 2 for document structure analyzes information regarding the constitution of the whole document which is inputted and sends the result to a sentence analytic part 3 together with the input sentence. The sentence analytic part 3 analyzes the input sentence and transfers part of the analytic result to a structuring process part 4, which converts the received analytic result into a sentence whose structure is clear and displays it at a display part 5. Over a look at this display, a user receives or reject the rewriting result selectively according to the display matches his or her thinking interpretation. Consequently, when the source is vague and possibly translated differently, and even when the user does not know the target language, the proper equivalent can be selected.



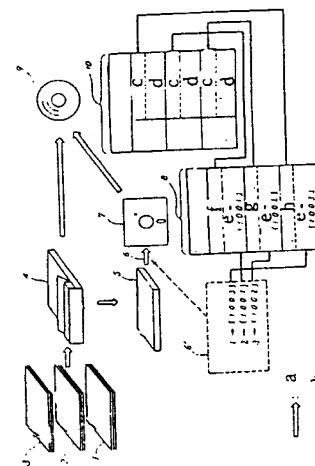
1: input document, 6: repetitive analytic result retrieval at sentence analytic part, a: sentence, b: document structure information

(54) AUTOMATIC RETRIEVAL SYSTEM FOR DOCUMENT CONTROL INFORMATION

(11) 4-153879 (A) (43) 27.5.1992 (19) JP
 (21) Appl. No. 2-280495 (22) 18.10.1990
 (71) FUJITSU LTD (72) YOSHIHIRO IKEDA
 (51) Int. Cl⁵. G06F15/40, G06F15/20, G06F15/62

PURPOSE: To automatically make a document and document control information correspond to each other by automatically retrieving document control information corresponding to a key value, read out of the contents of a document by a character recognition device when the document is registered, from a file on a different recording medium.

CONSTITUTION: When originals 1 - 3 of the document are inputted through a scanner 4, etc., and registered in a document file, the character recognition device 5 reads the key value out of the contents of the document. The document control information corresponding to this key value is retrieved from the document control information file 8 and registered in the document file 10 together with the originals of the document. Consequently, the serial operation wherein the document control information is added to the originals of the document and recorded in the document file can automatically be carried out and a user can easily make the originals and the contents of the document control information correspond to each other.



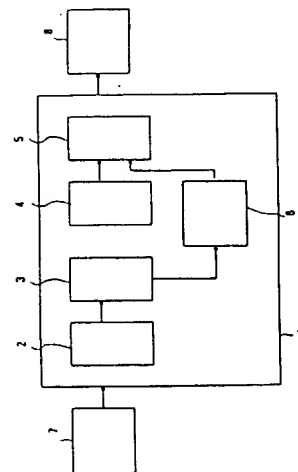
a: flow of information, b: correspondence, c: document control information part, d: original part, e: identification ID, f: document control information 1, g: document control information 2, h: document control information 3

(54) PARTIAL STRING EXTRACTION PROCESSING SYSTEM FOR CHARACTER STRING

(11) 4-153880 (A) (43) 27.5.1992 (19) JP
 (21) Appl. No. 2-280322 (22) 18.10.1990
 (71) PFU LTD (72) TADAHIRO MUROTANI
 (51) Int. Cl⁵. G06F15/40, G06F15/60

PURPOSE: To divide an input character string into specific partial strings and extract a partial string which meets specific requirements by storing attributes in an array type storage means in specific order by an attribute decision part and comparing the storage contents of the array type storage area with an extraction condition decision string by an extraction part.

CONSTITUTION: A division part 2 divides the input character string which is read by a reader 7 into character primitive strings according to predetermined conditions and the attribute decision part 3 decides the attributes of the respective divided character primitive strings by using predetermined conditions and stores the results in the array type storage area 6 in the specific order. An extraction condition generation part 4 generate the extraction condition decision string from extraction conditions read by the reader 7 and the extraction part 5 extracts the character primitive string corresponding to a string of matching attributes obtained by comparing the storage contents of the array type storage area 6, i.e. attributes with the extraction condition string. Consequently, only the character string matching the specific extraction requirements can automatically be extracted.



1: processor (CPU/memory), 3: output device

⑩ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A)

平4-153878

⑬ Int. Cl.⁵

G 06 F 15/38

識別記号

D
J

庁内整理番号

9194-5L
9194-5L

⑭ 公開 平成4年(1992)5月27日

審査請求 未請求 請求項の数 4 (全7頁)

⑮ 発明の名称 機械翻訳装置における前編集支援処理装置

⑯ 特 願 平2-280336

⑰ 出 願 平2(1990)10月18日

⑱ 発 明 者 伊 吹 潤 神奈川県川崎市中原区上小田中1015番地 富士通株式会社
内
⑱ 発 明 者 西 野 文 人 神奈川県川崎市中原区上小田中1015番地 富士通株式会社
内
⑱ 発 明 者 中 村 直 人 神奈川県川崎市中原区上小田中1015番地 富士通株式会社
内
⑱ 発 明 者 塩 内 正 利 神奈川県川崎市中原区上小田中1015番地 富士通株式会社
内
⑲ 出 願 人 富士通株式会社 神奈川県川崎市中原区上小田中1015番地
⑳ 代 理 人 弁理士 森 田 寛 外2名

最終頁に続く

明 細 書

1. 発明の名称

機械翻訳装置における前編集支援処理装置

2. 特許請求の範囲

- (1) 与えられた入力文書(1)について文章構造解析部(2)が文章の構造解析を行い、個々の文について文解析部(3)が単語の品詞決定を含む文解析を行って、上記与えられた入力文書(1)における原文を機械翻訳する機械翻訳装置において、

上記文解析部(3)によって行われた文解析の結果にもとづいて、原文の中の予め定められた文パターンについて、より構造が明確にされるように書き換える構造化処理部(4)と、

当該構造化処理部(4)において行われた結果の書き換え文を表示する表示部(5)とをそなえ、

上記より構造が明確にされた書き換え文を機

械翻訳処理に引継ぐようにした

ことを特徴とする機械翻訳装置における前編集支援処理装置。

- (2) 上記構造化処理部(4)は、原文の中の動詞の連用形と終助詞とをもつ文パターンについて、動詞の終止形と接続詞との形に変換するようにしたことを特徴とする請求項(1)記載の機械翻訳装置における前編集支援処理装置。

- (3) 上記構造化処理部(4)は、原文の中の特定の単語列についての評価値を出力する付与体(44)をそなえると共に、当該評価値にもとづいて上記原文を区分する可否かを指示する制御体(46)をそなえたことを特徴とする請求項(1)記載の機械翻訳装置における前編集支援処理装置。

- (4) 上記構造化処理部(4)は、原文における文構造を調べた上で原文を分割し、分割された夫々の文あるいはいずれか一方の文に対して文節を補完するようにしたことを特徴とする請求項(1)記載の機械翻訳装置における前編集支援処理装

置。

3. 発明の詳細な説明

〔概要〕

機械翻訳装置において利用者を支援する機械翻訳装置における前編集支援処理装置に関し、

原文があいまいさをもち、翻訳結果に複数の翻訳が得られる可能性がある場合などにおいて、利用者があいまいさのない形に原文を書き換え易くすることを目的とし、

機械翻訳装置における文解析部による解析結果について、予め定められたパターンの文を、より構造が明確にされるような形に書き換える構造化処理部をもうけ、当該構造化処理部による処理結果を表示するよう構成する。

〔産業上の利用分野〕

本発明は、機械翻訳装置において利用者を支援する機械翻訳装置における前編集支援処理装置に関する。

に読みにくいものになることが多かった。この場合、人間の作業者はシステムの出力から使えそうな部分訳をかき集め、自分の手で訳を構築するはめになり、作業効率の低下を招いていた。

即ち従来、機械翻訳装置から正しい翻訳を得るためには、利用者が原文をシステムが正しく解釈できるように形に書き換える必要があった。ところが、その作業はシステムの処理に精通した人が大変な手間をかけて行うものであり、機械翻訳を使った翻訳作業全体のコストを大きく引き上げていた。

この処理を自動的に行うシステムとして自動文分割システムなどが既に発表されているが、処理精度の点で問題が多く、実用的なものではなかった。

また訳文の読みやすさに対する配慮は、ほとんどなされておらず、このことも翻訳結果をチェックするのにかかる労力を引き上げ、ひいては翻訳にかかるコストの上昇する要因となっていた。

現在の機械翻訳装置では、意味的情報の欠如、パーズの記述力の不備などのため、文構造の解析を誤りなく行うことがむずかしい。このため、正しい翻訳結果を得るためには原文を曖昧さのない、一意に意味の定まるものとする処理（前編集）が必要となる。

〔従来の技術〕

現状では、当該前編集の処理に大きな手間がかかり、機械翻訳を利用した翻訳作業の能率の向上の妨げとなっている。これには次のような理由が考えられる。

- 1) エディタの操作性が悪く、書き換えに手間取る。
- 2) 翻訳システムの処理内容が明白に規定されていないため、原文をどのように修正すれば翻訳できるのかが判らない。

特に長文などの場合、前編集を行わない文の翻訳成功率は非常に低くなる。また翻訳が成功した場合でも、機械翻訳システムが出力する訳文は非常

〔発明が解決しようとする課題〕

初心者がシステムを使う場合、どのような形に翻訳対象文を修正したら正しい翻訳結果が得られるかは、容易にはわからない。そのため、修正なしでシステムにかけ、その結果、翻訳にかなりの誤りが混入するといった問題点があった。一方システムの処理ロジックに精通した人にとっても前編集作業は、文のほとんど全面的な書き換えを必要とするために極めて非能率的なものであった。さらに機械翻訳システムの翻訳結果は必ずしも読みやすいものではなく、翻訳結果を理解することや、翻訳結果を修正して最終的な翻訳を得ることを難しいものとしていた。

本発明は、原文があいまいさをもち、翻訳結果に複数の翻訳が得られる可能性がある場合などにおいて、利用者があいまいさのない形に原文を書き換え易くすることを目的としている。

〔課題を解決するための手段〕

第1図は本発明の原理構成図を示す。図中の符

号1は翻訳対象となる入力文書である。2は文章構造解析部であって、入力文書1を例えばタイトル、パラグラフ1、パラグラフ2、…の如く区分するなどの文章構造についての解析を行う。3は文解析部であって、入力文を解析する。4は構造化処理部であって、受けとった解析結果について構造が明確になるような形の文に変換する。5は表示部であって変換結果を表示する。6は文解析部3において得られている別の解析結果を求める別結果探索部を表わす。

本発明では、機械翻訳装置の使用にあたって、原文にあいまいさがある場合に、その可能な解析を原文を書き換えることによって示し、初心者が正しい翻訳が得られるような形に原文を修正する作業をシステムの示した書き換えの可能性の中から選択するだけで行えるように簡単化する。また文の構造がはっきりと示されるように文を書き換えることによって翻訳結果の読みやすさを改善するようにする。

書だけを目指す場合には、対話を行うための別結果探索部6を省略してもよい。

〔実施例〕

第2図は構造化処理部の一実施例構成を示し、第3図はその処理例を示す。第2図において符号4は構造化処理部であり、41は構成単語列解析体、42は構造化処理体を表わしている。

構成単語列解析体41は、入力文を構成する単語を解析し、各単語に対応した品詞を判定する。構造化処理体42は予め定められた文パターンを抽出し、当該文パターンをもつ文をより明確な文に変換する。

第3図図示の処理例は、

動詞の連用形+終助詞「て」

で与えられる文パターンをもつ文について、

動詞の終止形+「。」+「そして」

の形に変換する例を表わしている。「処理例1」の場合における「消して」はサ行5段活用をとっているものであり、「ファイルを消して処理

〔作用〕

文章構造の解析部2は、入力された文章全体の構成に関する情報を解析して、入力文とともに、文解析部3に送る。文解析部3は、入力文を解析してその解析結果の一部を構造化処理部4に転送する。構造化処理部4は、受けとった解析結果を構造がはっきりと判るような形の文へ変換して表示部5に表示する。当該表示をみてユーザは自分の考えた解釈と一致するか否かによって書き換え結果を受理するか、拒否するかを選択する。

拒否された場合には、システムは更に別の解析結果を文解析部3から転送し、再び、構造化処理部4の処理にゆだねる。受理する場合には、解析結果はそのまま出力される。この時点ではあいまいな部分についてどのような解釈がとられているかが明らかにされ、ユーザに承認されていることとなる。

もちろん、上記の場合は、ユーザとの対話によって翻訳の曖昧さを解消することを目的とした処理の場合である。単に翻訳結果の読みやすさの改

を終える。」という入力文は、「ファイルを消す。そして処理を終える。」という形に変換される。「処理例2」の場合における「消去して」はサ行変格活用をとっているものであり、「ファイルを消去して処理を終える。」という入力文は、「ファイルを消去する。そして処理を終える。」という形に変換される。

この処理の場合、例えば「AしてBしてCする。」の如く連なる場合には、「Aする。そしてBする。そしてCする。」の如く変換されると考えてよい。

第4図は構造化処理部の他の実施例構成を示し、第5図はその処理例を示す。第4図において符号4は構造化処理部であり、43は形態素解析体、44は特定単語列の走査と評価値の付与体（以下、評価値付与体と略す）、45は構造化処理体、46は構造化処理の制御体を表わしている。

第4図図示の場合には、得られた単語列にもとづいて、評価値付与体44が、例えば述語が3個以上存在する如き文を調べる。そして構造化処理

部4は、当該3個以上の場合に限り、述語が2個以内の文の形に変換するようにする。即ち、評価値付与体44が評価値「述語の個数」を制御体46に与え、当該制御体46が「述語3個以上」の条件に合致するか否かをチェックし、構造化処理体45に対して、原文を変換するか否かの指示「on/offの制御信号」を与える。

第5図図示の処理例は、「述語が3つ以上存在する」場合に「文の分割化処理を行う」ようにした例を表わしている。

図示(i)の場合には、「ファイルを消去して処理を終える。」という文において、述語が2個であることから、文の分割化処理は行われない。それに対して、図示(ii)の場合には、「データを退避し、ファイルを消去して処理を終える。」という文において、述語が3個であることから、文の分割化処理が行われる。この例の場合には、「退避し」がサ行変格活用であり、「消去し」がサ行変格活用であることから、「データを退避し、ファイルを消去する。そして処理を終える。」という形に

7図(B)における「ユーザは、」の代わりに第7図(C)においては「彼は、」を補なうようにしている。

したがって、第8図に処理例を示す如く、

「ユーザは、ファイルを消去して処理を終える。」

という原文は、

「ユーザは、ファイルを消去する。」

「そして彼は、処理を終える。」

の如く変換されることとなる。第9図図示の場合には第8図図示の場合と同様に変換されるが、併せて単語の品詞をも表示するようにした場合を表わしている。

(発明の効果)

以上説明した如く、本発明によれば、原文中にあいまいさがあり、複数の翻訳の可能性がある場合、かつユーザが目標言語を知らない場合においても適切な訳を選択することが可能となる。また本発明によれば、ユーザが機械翻訳システムの処

変換される。

第6図は構造化処理部の更に他の実施例を示し、第7図は文の構造情報を利用した構造化の例、第8図および第9図は夫々処理結果の例を示す。

第6図において、符号4は構造化処理部であり、43は形態素解析体、47は文構造解析体、48は構造化処理体本体を表わしている。

第6図図示の場合には、第7図を参照して説明する如く、文構造解析体47が文中の階層関係を明らかにし構造化処理体本体48が原文を変換するようにしている。

第7図(A)は原文の「ユーザは、ファイルを消去して処理を終える。」について文構造を明らかにした状態を表わしている。図中のsは文、vpは述語、ppは文節を表わしている。第7図(A)においては、(i)「ファイルを消去して」の部分と、(ii)「処理を終える。」の部分とが、同じ「ユーザは、」の述語に対応している。この結果、第7図(B)と第7図(C)との如く、2つの文に分割し、第7図(C)の場合の如く、第

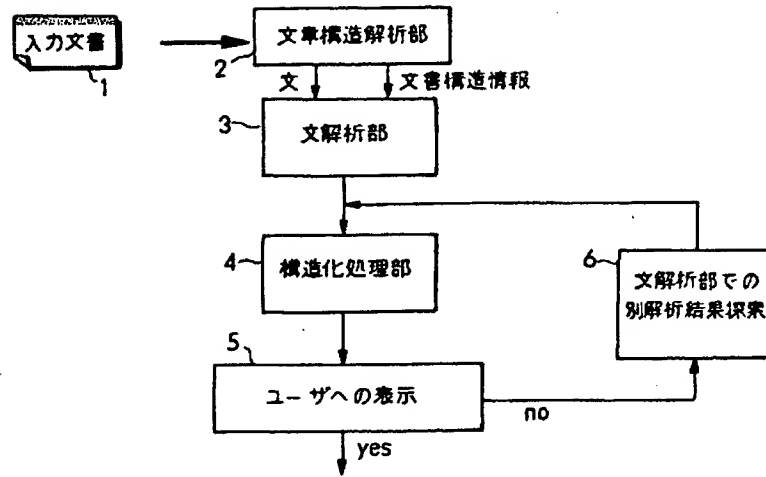
理できる範囲の文についての知識がなく、目標言語を知らない場合においても適切な訳を選択することが可能となる。

なお本発明においては、上記第2図、第4図、第6図図示の各構成のいずれか1つのみの場合であってよいことは言うまでもなく、いずれか複数個の組合わせをもつ場合であってもよい。

4. 図面の簡単な説明

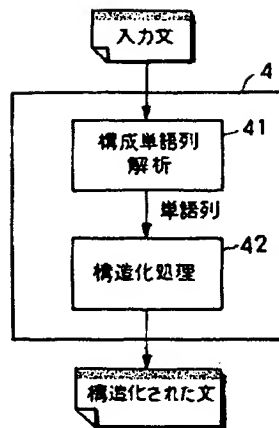
第1図は本発明の原理構成図、第2図は構造化処理部の一実施例、第3図は第2図図示の場合の処理例、第4図は構造化処理部の他の実施例、第5図は第4図図示の場合の処理例、第6図は構造化処理部の更に他の実施例、第7図は文の構造情報を利用した構造化の例、第8図および第9図は夫々処理結果の例を示す。

図中、1は入力文書、2は文章構造解析部、3は文解析部、4は構造化処理部、5は表示部、6は別結果探索部を表わしている。



本発明の原理構成図

第 1 図



構造化処理部

第 2 図

処理の概要
 動詞の連用形 + 終助詞「て」
 →
 動詞の終止形 + 「。」 + 「でして」

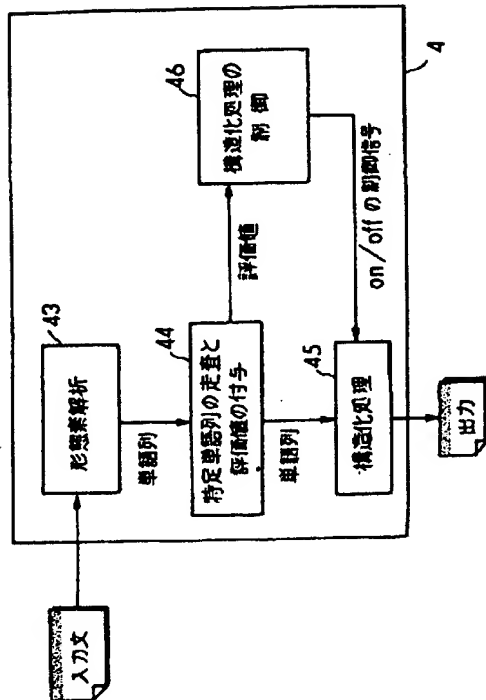
処理例 1
 ファイルを消して処理を終える。
 ファイル 名詞
 を 格助詞
 消し 他動詞 サ行5段活用 連用形
 て 終助詞
 処理 名詞
 を 格助詞
 終える 他動詞
 。 記号

→
 ファイルを消す。でして処理を終える。

処理例 2
 ファイルを消去して処理を終える。
 ファイル 名詞
 を 格助詞
 消去し 他動詞 サ行変格活用 連用形
 て 終助詞
 処理 名詞
 を 格助詞
 終える 他動詞 終止形
 。 記号

→
 ファイルを消去する。でして処理を終える。

第 3 図



構造化処理部
第4図

処理の概要

述語が3つ以上存在する。

文の分割化処理を行なう。

(1) ファイルを消して処理を終える。(述語数2)

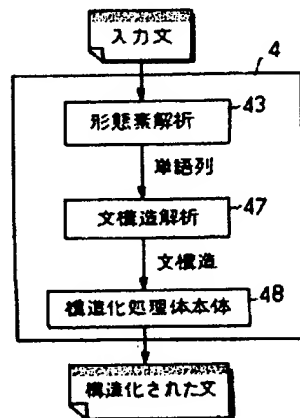
ファイルを消して処理を終える。(そのまま)

(2) データを通過し、ファイルを消去して処理を終える。(述語数3)

データ	名詞
を	格助詞
通過し	他動詞 2行文格活用 適用形
,	記号
ファイル	名詞
を	格助詞
消去し	他動詞 2行文格活用 適用形
て	終助詞
処理	名詞
を	格助詞
終える	他動詞 終止形
.	記号

データを通過し、ファイルを消去する。そして処理を終える。

処理例
第5図



構造化処理部
第6図

処理例

ユーザは、ファイルを消去して処理を終える。

ユーザは、ファイルを消去する。
そして彼は、処理を終える。

最終結果として文字列を出力する例
第8図

処理例

ユーザは、ファイルを消去して処理を終える。

ユーザは、ファイルを消去する。

ユーザ	名詞
は	格助詞
,	記号
ファイル	名詞
を	格助詞
消去する	他動詞
.	記号

そして彼は、処理を終える。

そして	接続詞
彼	名詞
は	格助詞
,	記号
処理	名詞
を	格助詞
終える	他動詞
.	記号

最終結果として単語列を出力する例
第9図

